

Superconcentration and related phenomena

Sourav Chatterjee

Courant Institute, NYU

Variance of a random variable

- ▶ If X is a random variable with expected value $\mathbb{E}(X)$, then the variance of X is defined as

$$\text{Var}(X) := \mathbb{E}(X - \mathbb{E}(X))^2.$$

Square-root of the variance is a measure of the “average size of the fluctuations of X ”.

- ▶ The variance of a sum is the sum of covariances. Gives an easy way to compute variances in many classical examples such as sums of independent or weakly dependent random variables, U -statistics, subgraph counts in random graphs, etc.
- ▶ Less ambitious than computing exponential tail bounds. **Yet, modern probability theory is not powerful enough to estimate variances in truly challenging problems.** This is the focus of this lecture series.

Challenging example # 1: First-passage and last-passage percolation

- ▶ Let $G = (V, E)$ be an undirected graph; let each edge $e \in E$ carry a weight ω_e , where the weights are i.i.d. random variables.
- ▶ If P is a self-avoiding path in G , let $\omega(P)$ denote the sum of the edge-weights along the path.
- ▶ Let \mathcal{P} be a collection of self-avoiding paths in G .
- ▶ Let $L := \max\{\omega(P) : P \in \mathcal{P}\}$ be the **last-passage time** for this set of paths. Similarly, $T := \min\{\omega(P) : P \in \mathcal{P}\}$ is called the **first-passage time**.
- ▶ Other than in very special exactly solvable models, we do not know how to compute or even estimate the correct order of the variance of L .

Challenging example #1 (contd.)

- ▶ Special case: $G = \mathbb{Z}^d$. \mathcal{P} is the set of **all positively oriented paths** from the origin to (N, \dots, N) . Last-passage time denoted by L_N .
- ▶ Similarly, with $G = \mathbb{Z}^d$ and \mathcal{P} the set of **all paths** from the origin to $(N, 0, 0, \dots, 0)$, the first-passage time is denoted by T_N . Only makes sense when the weights are non-negative.
- ▶ For T_N with binary edge-weights, Benjamini, Kalai and Schramm showed that the variance is bounded by $CN/\log N$, improving a bound of Kesten. Extended to a larger class of edge-weights by Benaïm and Rossignol.
- ▶ For L_N with exponential vertex-weights in $d = 2$, the variance is of order $N^{2/3}$ (follows from exact solution by Johansson). For Gaussian weights, I showed that $\text{Var}(L_N) \leq CN/\log N$ in $d = 2$, which was generalized to all dimensions by Graham.
- ▶ Lastly, a lot is known about last-passage percolation on trees (Bramson, Zeitouni, etc.).

Challenging example # 2: Spin glass free energy

- ▶ N particles. Particle i has spin $\sigma_i \in \{-1, 1\}$.
- ▶ Configuration: $\sigma = (\sigma_1, \dots, \sigma_N)$.
- ▶ $H_N(\sigma) =$ energy of the configuration σ .
- ▶ In spin glasses, the energy function is not deterministic; $\{H_N(\sigma) : \sigma \in \{-1, 1\}^N\}$ is modeled as a collection of random variables. This is sometimes called the **energy landscape**.
- ▶ Mean-field spin glasses: $H_N(\sigma)$ are jointly Gaussian random variables, with zero mean and

$$\text{Cov}(H_N(\sigma), H_N(\sigma')) = N\psi\left(\frac{\sigma \cdot \sigma'}{N}\right),$$

where ψ is a given function on $[-1, 1]$.

- ▶ Example: **Sherrington-Kirkpatrick (S-K) model**. $\psi(x) = \frac{1}{2}x^2$.

Challenging example # 2 (contd.)

- ▶ In the S-K model, an alternative way to represent the energy is

$$H_N(\sigma) = -\frac{1}{\sqrt{2N}} \sum_{i,j=1}^N g_{ij} \sigma_i \sigma_j$$

where g_{ij} are i.i.d. standard Gaussian random variables.

- ▶ The collection $(g_{ij})_{1 \leq i,j \leq N}$ is collectively called the **disorder**.
- ▶ **Gibbs measure**: random probability measure on $\{-1, 1\}^N$ that puts mass

$$Z_N(\beta)^{-1} e^{-\beta H_N(\sigma)}$$

on the configuration σ . Here β is a parameter called the **inverse temperature** and $Z_N(\beta)$ is the normalizing constant.

Challenging example # 2 (contd.)

- ▶ **Free energy:** $F_N(\beta) = \frac{1}{\beta} \log Z_N(\beta)$.
- ▶ Question: What is the variance of $F_N(\beta)$?
- ▶ Other than in very special situations, we do not know how to compute the exact order of the variance of the free energy in spin glass models (as a function of N).
- ▶ Aizenman, Lebowitz and Ruelle showed that the variance is $O(1)$ if $\beta < 1$. This is optimal.
- ▶ For $\beta > 1$, standard techniques show that $\text{Var}(F_N(\beta)) \leq C(\beta)N$. I improved this to $C(\beta)N/\log N$.
- ▶ Again, these are the best known results.

- ▶ Suppose that $(X^t)_{t \geq 0}$ is a stationary, reversible Markov process with invariant measure μ . The semigroup $(P_t)_{t \geq 0}$ induced by this process is defined as

$$P_t f(x) = \mathbb{E}(f(X^t) \mid X^0 = x).$$

- ▶ The generator L for the process is

$$Lf(x) = \lim_{t \rightarrow 0} \frac{P_t f(x) - f(x)}{t}.$$

- ▶ Heat equation: $\partial_t P_t f = L P_t f$. Consequently, $P_t = e^{tL}$.
- ▶ Inner product: $(f, g) = \int f(x)g(x)d\mu(x)$.
- ▶ L is self-adjoint: $(f, Lg) = (Lf, g)$.
- ▶ Dirichlet form: $\mathcal{E}(f, g) := -(f, Lg)$.

Semigroup basics (contd.)

- ▶ Often, $-L$ is a positive semidefinite operator with a countable sequence of eigenvalues $0 = \lambda_0 < \lambda_1 \leq \dots$ and a corresponding sequence of orthonormal eigenvectors u_0, u_1, \dots . The assumption that $\lambda_1 > 0$ is same as saying that P_t has a **spectral gap**.
- ▶ Therefore,

$$P_t f = e^{tL} f = \sum_{i=0}^{\infty} e^{-\lambda_i t} (u_i, f) u_i, \quad \text{and}$$

$$\mathcal{E}(f, g) = \sum_{i=0}^{\infty} \lambda_i (u_i, f) (u_i, g).$$

- ▶ As a consequence, $\mathcal{E}(f, f) \geq 0$ for any f . Moreover,

$$\mathcal{E}(f, P_t f) = \sum_{i=1}^{\infty} \lambda_i e^{-\lambda_i t} (u_i, f)^2,$$

a formula that we will use later.

The classical attack on the variance: Poincaré inequalities

- ▶ Suppose that we have a random object X defined on some space, and $Y = f(X)$ is a real-valued function of X .
- ▶ The Poincaré inequality approach to bounding the variance of Y goes as follows:
 - ▶ Construct a stationary reversible Markov process X^t with $X^0 = X$.
 - ▶ Show that the Markov process satisfies a **Poincaré inequality**, that is, there exists a constant C such that for any function g ,

$$\text{Var}(g(X)) \leq C\mathcal{E}(g, g).$$

- ▶ Apply the Poincaré inequality to get the bound

$$\text{Var}(Y) \leq C\mathcal{E}(f, f).$$

Usually, the right-hand side is easy to bound.

Example: The Efron-Stein inequality

- ▶ Suppose that X_1, \dots, X_n are independent random variables and $Y = f(X_1, \dots, X_n)$.
- ▶ Here $X = (X_1, \dots, X_n)$.
- ▶ Stationary Markov process: Take n independent Poisson clocks. Whenever clock i rings, replace the i th coordinate by an independent pick from the distribution of X_i . This may be called the **Poisson flipping dynamics**.
- ▶ One can show that this process satisfies a Poincaré inequality with constant n .
- ▶ This gives the so-called Efron-Stein inequality:

$$\begin{aligned} & \text{Var}(f(X_1, \dots, X_n)) \\ & \leq \frac{1}{2} \sum_{i=1}^n \mathbb{E}(f(X_1, \dots, X_n) - f(X_1, \dots, X'_i, \dots, X_n))^2, \end{aligned}$$

where X'_i is an independent copy of X_i .

Another example: The Gaussian Poincaré inequality

- ▶ In many examples such as the S-K model, X is a d -dimensional standard Gaussian random vector.
- ▶ The relevant Markov process is the d -dimensional Ornstein-Uhlenbeck process, described by the SDE $dX^t = -X^t dt + \sqrt{2}dB^t$. The Dirichlet form for the process is

$$\mathcal{E}(f, g) = \int \nabla f \cdot \nabla g \, d\gamma,$$

where γ is the d -dimensional standard Gaussian measure.

- ▶ The standard Gaussian measure is the unique invariant measure for this process.
- ▶ Moreover, the process satisfies a Poincaré inequality with constant 1, irrespective of the dimension.
- ▶ The Poincaré inequality takes the form

$$\text{Var}(f(X)) \leq \mathbb{E}|\nabla f(X)|^2.$$

This is known as the **Gaussian Poincaré inequality**.

What do Poincaré inequalities give in our examples?

- ▶ In oriented last-passage percolation from $(0, \dots, 0)$ to (N, \dots, N) , the Efron-Stein inequality gives $\text{Var}(L_N) \leq CN$, where C is a constant depending on the distribution of the edge-weights.
- ▶ In the Sherrington-Kirkpatrick model, the Gaussian Poincaré inequality gives $\text{Var}(F_N(\beta)) \leq C(\beta)N$.
- ▶ These are essentially the best bounds one can get by classical methods. Can be refined to give sub-Gaussian tail bounds, but the bounds on the order of fluctuations do not improve.
- ▶ The bounds are, however, not optimal.

The semigroup approach to Poincaré inequalities

- ▶ Poincaré inequalities are often proved using the **semigroup method**.
- ▶ First step:

$$\begin{aligned}\mathrm{Var}_\mu(f) &= (f, P_0 f) - (f, P_\infty f) \\ &= - \int_0^\infty \partial_t (f, P_t f) dt \\ &= - \int_0^\infty (f, \partial_t P_t f) dt \\ &= - \int_0^\infty (f, L P_t f) dt = \int_0^\infty \mathcal{E}(f, P_t f) dt.\end{aligned}$$

- ▶ Second step: Using special properties of the semigroup, prove that $\mathcal{E}(f, P_t f) \leq e^{-ct} \mathcal{E}(f, f)$.

The Fourier analytic approach to Poincaré inequalities

- ▶ Recall the eigenvalues λ_i and eigenvectors u_i of the operator $-L$. Then

$$\text{Var}_\mu(f) = \sum_{i=1}^{\infty} (u_i, f)^2, \quad \text{and}$$
$$\mathcal{E}(f, f) = \sum_{i=1}^{\infty} \lambda_i (u_i, f)^2,$$

which automatically prove the Poincaré inequality

$$\text{Var}_\mu(f) \leq \frac{1}{\lambda_1} \mathcal{E}(f, f).$$

- ▶ In fact, $1/\lambda_1$ is the optimal constant.

An improvement of the Poincaré inequality by the Fourier analytic approach

- ▶ Since $\lambda_1 \leq \lambda_2 \leq \dots$, for any $k \geq 1$, we have

$$\begin{aligned}\mathrm{Var}_\mu(f) &= \sum_{i=1}^{k-1} (u_i, f)^2 + \sum_{i=k}^{\infty} (u_i, f)^2 \\ &\leq \sum_{i=1}^{k-1} (u_i, f)^2 + \frac{1}{\lambda_k} \sum_{i=k}^{\infty} \lambda_i (u_i, f)^2 \\ &\leq \sum_{i=1}^{k-1} (u_i, f)^2 + \frac{1}{\lambda_k} \mathcal{E}(f, f).\end{aligned}$$

- ▶ For the O-U semigroup, the d -dimensional Hermite polynomials provide a suitable orthonormal basis for the Hilbert space $L^2(\gamma)$.
- ▶ They are a collection $(H_k : k \in \mathbb{Z}_+^d)$ of polynomials in d variables.
- ▶ For $k = (k_1, \dots, k_d)$, the eigenvalue corresponding to H_k is simply $k_1 + \dots + k_d$.

Fourier analysis for the O-U semigroup (contd.)

- ▶ Using integration-by-parts, one can show that if $k = (k_1, \dots, k_d)$, then

$$(f, H_k) = \frac{1}{\sqrt{k_1! \cdots k_d!}} \int \frac{\partial^{k_1 + \cdots + k_d} f}{\partial x_1^{k_1} \cdots \partial x_d^{k_d}} d\gamma.$$

- ▶ From this it is easy to show that if we define

$$\theta_m(f) := \sum_{k \in \mathbb{Z}_+^d : k_1 + \cdots + k_d = m} (f, H_k)^2,$$

then

$$\theta_m(f) = \frac{1}{m!} \sum_{1 \leq i_1, \dots, i_m \leq d} \left(\int \frac{\partial^m f}{\partial x_{i_1} \cdots \partial x_{i_m}} d\gamma \right)^2.$$

An improvement of the Gaussian Poincaré inequality

- ▶ For the O-U semigroup, our previous Fourier analytic improvement of the Poincaré inequality takes the following form: For any $M \geq 1$,

$$\mathrm{Var}_\gamma(f) \leq \sum_{m=1}^{M-1} \theta_m(f) + \frac{1}{M} \mathbb{E}_\gamma |\nabla f|^2.$$

- ▶ For $M = 1$, this is the usual Poincaré inequality.

Implication for the S-K model

- ▶ Recall: The free energy $F_N(\beta)$ of the S-K model is a function of the disorder $(g_{ij})_{1 \leq i, j \leq N}$.
- ▶ When $\beta > 1$, one can show that

$$c(\beta)N \leq \mathbb{E}|\nabla F_N(\beta)|^2 \leq C(\beta)N.$$

- ▶ However, it is not too difficult to prove the crude upper bound

$$\theta_m(F_N(\beta)) \leq (C(\beta)m)^{C(\beta)m}$$

for each $m \geq 1$. Note that this bound does not depend on N , but blows up as $m \rightarrow \infty$ (whereas the true θ_m must go to zero).

- ▶ Choosing $M = c \log N / \log \log N$ for some small enough c and applying the improved Poincaré inequality gives

$$\text{Var}(F_N(\beta)) \leq \frac{C(\beta)N \log \log N}{\log N}.$$

- ▶ This best available bound is $N / \log N$ (C., 2009).

Superconcentration

- ▶ Recall: The Poincaré inequality for the d -dimensional standard Gaussian measure γ says that $\text{Var}_\gamma(f) \leq \mathbb{E}_\gamma |\nabla f|^2$.
- ▶ We will say that f is **superconcentrated** with respect to γ if $\text{Var}_\gamma(f)$ is very small compared to $\mathbb{E}_\gamma |\nabla f|^2$. (For example, the free energy of the S-K model when N is large.)
- ▶ More precisely, we will say that f is **ϵ -superconcentrated** with respect to γ if

$$\text{Var}_\gamma(f) \leq \epsilon \mathbb{E}_\gamma |\nabla f|^2.$$

- ▶ When $\beta > 1$, we saw that the free energy $F_N(\beta)$ of the S-K model is ϵ_N -superconcentrated, where $\epsilon_N = C(\beta) \log \log N / \log N$. Since $\epsilon_N \rightarrow 0$ as $N \rightarrow \infty$, we will simply say that $F_N(\beta)$ is superconcentrated. (When $\beta \leq 1$, the free energy is not superconcentrated.)
- ▶ We will see later that superconcentration has important implications about the structure of the problem.

Superconcentration (contd.)

- ▶ The definition of superconcentration can be extended beyond the Gaussian measure.
- ▶ Given any Markov process with invariant measure μ and Dirichlet form \mathcal{E} that satisfies a Poincaré inequality with optimal constant C , we will say that a function f is ϵ -superconcentrated with respect to the Markov process (or, by abuse of terminology, with respect to μ), if

$$\text{Var}_\mu(f) \leq C \epsilon \mathcal{E}(f, f).$$

- ▶ Since $\text{Var}_\mu(f) = \sum_{i=1}^{\infty} (u_i, f)^2$ and $\mathcal{E}(f, f) = \sum_{i=1}^{\infty} \lambda_i (u_i, f)^2$, **superconcentration occurs if and only if most of the “Fourier mass” concentrates on the higher end of the spectrum.**

Hypercontractivity

- ▶ A semigroup P_t is said to be **hypercontractive** if for all $p \geq 1$ and $t > 0$, there is some $q(t, p) > p$ such that for all f ,

$$\|P_t f\|_{q(t,p)} \leq \|f\|_p.$$

- ▶ For example, the O-U semigroup is hypercontractive with $q(t, p) = 1 + (p - 1)e^{2t}$.
- ▶ When $X = (X_1, \dots, X_n)$ is a vector of i.i.d. **binary** random variables, the Poisson flipping dynamics is hypercontractive.
- ▶ Usually proved using log-Sobolev inequalities (Gross's method).

The L^1 - L^2 method

- ▶ The first general method for proving superconcentration was developed by Talagrand using a clever mix of hypercontractivity and the semigroup method.
- ▶ Given a function $f : \{0, 1\}^n \rightarrow \mathbb{R}$, define

$$\begin{aligned}\Delta_i f(x_1, \dots, x_n) &:= f(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_n) \\ &\quad - f(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n).\end{aligned}$$

- ▶ Let μ be the uniform distribution on $\{0, 1\}^n$. Then Talagrand proved that

$$\text{Var}_\mu(f) \leq C \sum_{i=1}^n \frac{\|\Delta_i f\|_{L^2(\mu)}^2}{1 + \log \frac{\|\Delta_i f\|_{L^2(\mu)}}{\|\Delta_i f\|_{L^1(\mu)}}},$$

where C is a universal constant.

- ▶ Compare with the Efron-Stein inequality, which is the same but without the log factor.

The L^1 - L^2 method (contd.)

- ▶ One can similarly write down an L^1 - L^2 upper bound for the d -dimensional Gaussian measure:

$$\text{Var}_\gamma(f) \leq C \sum_{i=1}^d \frac{\|\partial_i f\|_{L^2(\gamma)}^2}{1 + \log \frac{\|\partial_i f\|_{L^2(\gamma)}}{\|\partial_i f\|_{L^1(\gamma)}}}.$$

- ▶ Sketch of proof: Semigroup method gives

$\text{Var}_\gamma(f) = \int_0^\infty \mathbb{E}_\gamma(\nabla f \cdot \nabla P_t f) dt$. For the O-U semigroup, $\nabla P_t f = e^{-t} P_t \nabla f$. Therefore,

$$\text{Var}_\gamma(f) = \int_0^\infty e^{-t} \sum_{i=1}^d (\partial_i f, P_t \partial_i f) dt.$$

By Cauchy-Schwarz, $|(\partial_i f, P_t \partial_i f)| \leq \|\partial_i f\|_{L^2(\gamma)} \|P_t \partial_i f\|_{L^2(\gamma)}$.
By hypercontractivity, $\|P_t \partial_i f\|_{L^2(\gamma)} \leq \|\partial_i f\|_{L^{p(t)}(\gamma)}$, where $p(t) \in (1, 2)$. The rest is analytic manipulation.

A simple example

- ▶ Let $f(x_1, \dots, x_d) = \max x_i$. Then

$$\partial_i f = 1_{\{x_i \geq x_j \text{ for all } j\}}.$$

- ▶ Let γ be the d -dimensional standard Gaussian measure. Then

$$\|\partial_i f\|_{L^2(\gamma)}^2 = \|\partial_i f\|_{L^1(\gamma)} = \mathbb{P}(X_i \geq X_j \text{ for all } j) = \frac{1}{d}.$$

- ▶ Usual Poincaré inequality gives

$$\text{Var}_\gamma(f) \leq \sum_{i=1}^d \|\partial_i f\|_{L^2(\gamma)}^2 = 1.$$

- ▶ On the other hand, the L^1 - L^2 bound gives

$$\text{Var}_\gamma(f) \leq \frac{C}{\log d},$$

which is of the correct order.

Applications of the L^1 - L^2 method

- ▶ The first major application of the L^1 - L^2 method was by Benjamini, Kalai and Schramm, who improved the classical upper bound on the variance of the first-passage percolation time (with binary edge-weights) to $N/\log N$.
- ▶ Extended to other edge-weights by Benaïm and Rossignol.
- ▶ I used the BKS method to similarly improve the upper bound on last-passage percolation time in $d = 2$. Extended to $d \geq 3$ by Graham.
- ▶ Several other uses in recent years.

A variant of the L^1 - L^2 bound for the Gaussian measure

- ▶ Let $v_i = \|\partial_i f\|_{L^1(\gamma)}^2$ and $w_i = \|\partial_i f\|_{L^2(\gamma)}^2$, so that

$$\text{Var}_\gamma(f) \leq C \sum_{i=1}^d \frac{w_i}{1 - \frac{1}{2} \log \frac{v_i}{w_i}} = C \sum_{i=1}^d g(v_i/w_i) w_i,$$

where $g(x) = 1/(1 - \frac{1}{2} \log x)$.

- ▶ The function g is concave on $(0, 1]$. So by Jensen's inequality,

$$\text{Var}_\gamma(f) \leq C g\left(\frac{\sum v_i}{\sum w_i}\right) \sum_{i=1}^d w_i$$

- ▶ Thus, if we define

$$B(f) := \frac{\sum \|\partial_i f\|_{L^1(\gamma)}^2}{\sum \|\partial_i f\|_{L^2(\gamma)}^2},$$

then f is ϵ -superconcentrated, where $\epsilon = C/(1 - \frac{1}{2} \log B(f))$.

The L^1 - L^2 method gives a necessary and sufficient criterion for superconcentration of monotone functions

- ▶ Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be monotone in each coordinate.
- ▶ The Hermite polynomial expansion gives

$$\text{Var}_\gamma(f) = \sum_{m=1}^{\infty} \theta_m(f) \geq \theta_1(f) = \sum_{i=1}^d (\mathbb{E}_\gamma \partial_i f)^2.$$

- ▶ Since $\partial_i f \geq 0$ everywhere, $\mathbb{E}_\gamma \partial_i f = \|\partial_i f\|_{L^1(\gamma)}$.
- ▶ Thus,

$$\text{Var}_\gamma(f) \geq B(f) \mathbb{E}_\gamma |\nabla f|^2.$$

- ▶ Combining this bound with the previous one, we get

$$B(f) \leq \frac{\text{Var}_\gamma(f)}{\mathbb{E}_\gamma |\nabla f|^2} \leq \frac{C}{1 - \frac{1}{2} \log B(f)}.$$

Thus, f is superconcentrated if and only if $B(f)$ is small. (C., 2012)

Superconcentration in last-passage percolation

- ▶ Recall the generalized last-passage percolation problem: There is a graph $G = (V, E)$, and a set of self-avoiding paths \mathcal{P} in G . Each edge e has a weight ω_e attached to it. For a path P , $\omega(P)$ is the sum of edges weights along P . The last-passage time L is defined as

$$L = \max_{P \in \mathcal{P}} \omega(P).$$

- ▶ Assume that the ω_e 's are i.i.d. standard Gaussian random variables.
- ▶ Note that $\partial_e L = 1\{e \in \text{optimal path}\}$, where $\partial_e L$ is the derivative of L with respect to ω_e . Moreover, L is a monotone function.
- ▶ Consequently, if $p_e := \mathbb{P}(e \in \text{optimal path})$, then

$$B(L) = \frac{\sum p_e^2}{\sum p_e}.$$

Superconcentration in last-passage percolation (contd.)

- ▶ Let \hat{P} be the optimal path and let $|\hat{P}|$ be the number of edges in the optimal path.
- ▶ Let \hat{P}' be the optimal path in a different environment ω' , where ω' is independent of ω and has the same distribution.
- ▶ Note that $\sum p_e = \mathbb{E}|\hat{P}|$ and $\sum p_e^2 = \mathbb{E}|\hat{P} \cap \hat{P}'|$.
- ▶ Thus, L is superconcentrated if and only if $\mathbb{E}|\hat{P} \cap \hat{P}'|$ is negligible compared to $\mathbb{E}|\hat{P}|$. **This gives a necessary and sufficient condition for superconcentration in last-passage percolation on arbitrary graphs.** (C., 2012)

Some counterexamples

“Taking the maximum of a lot of almost independent Gaussian variables do not automatically give superconcentration.”

- ▶ For any $\sigma = (\sigma_1, \dots, \sigma_n) \in \{-1, 1\}^n$, let

$$X_\sigma := \frac{1}{\sqrt{n}} \sum_{i=1}^n g_i \sigma_i,$$

where g_1, \dots, g_n are i.i.d. standard Gaussian random variables.

- ▶ Then each X_σ is a standard Gaussian r.v. Moreover, for typical σ and σ' , the correlation between X_σ and $X_{\sigma'}$ is close to zero.
- ▶ However, if we let $M := \max X_\sigma$, then

$$M = \frac{1}{\sqrt{n}} \sum_{i=1}^n |g_i|,$$

which has variance bounded below by a constant.

- ▶ The above expression also implies that $\mathbb{E}_\gamma |\nabla M|^2 = O(1)$, and so M is **not superconcentrated**.

Some counterexamples (contd.)

“Taking the maximum of a larger collection of Gaussian variables does not always make the variance smaller, even if the variables are positive correlated and have the same variance.”

- ▶ Let g_{ij} , $i, j = 1, \dots, n$ be i.i.d. standard Gaussian r.v.
- ▶ Let Π be the set of all maps from $\{1, 2, \dots, n\}$ into itself.
- ▶ For any $\pi \in \Pi$, let

$$X_\pi := \sum_{i=1}^n g_{i\pi(i)}.$$

- ▶ Let $M := \max_{\pi \in \Pi} X_\pi = \sum_{i=1}^n \max\{g_{i1}, \dots, g_{in}\}$.
- ▶ $\text{Var}(M)$ is of order $n/\log n$, since each term in the above sum has variance of order $1/\log n$ and the terms are independent.
- ▶ Given $A \subseteq \Pi$, let $M_A := \max_{\pi \in A} X_\pi$. We will show that there is a subset A such that $\text{Var}(M_A) = O(n^{1/2})$.

Some counterexamples (contd.)

- ▶ Let A be a subset of Π of size $e^{cn^{1/2}}$ such that for all $\pi, \pi' \in A$, $\pi \neq \pi'$, we have

$$|\pi \cap \pi'| := |\{i : \pi(i) = \pi'(i)\}| \leq n^{1/2}.$$

Such an A is easily found by choosing $e^{cn^{1/2}}$ elements of Π uniformly at random, and then proving that the above event has positive probability.

- ▶ Let $\partial_{ij} M_A$ denote the derivative of M_A with respect to g_{ij} . Then

$$\partial_{ij} M_A = 1\{\hat{\pi}(i) = j\},$$

where $\hat{\pi}$ is the optimal map in A .

- ▶ Now suppose that the array (g_{ij}^t) is flowing in time as an O-U flow, with $g_{ij}^0 = g_{ij}$. Let $\hat{\pi}^t$ denote the optimal map at time t . The semigroup method gives

$$\text{Var}(M_A) = \int_0^\infty e^{-t} \mathbb{E} |\hat{\pi} \cap \hat{\pi}^t| dt.$$

Some counterexamples (contd.)

- ▶ But we know that either $\hat{\pi} = \hat{\pi}^t$, or $|\hat{\pi} \cap \hat{\pi}^t| \leq n^{1/2}$.
- ▶ Thus,

$$\text{Var}(M_A) \leq n^{1/2} + n \int_0^\infty e^{-t} \mathbb{P}(\hat{\pi} = \hat{\pi}^t) dt.$$

- ▶ Using a mix of hypercontractivity, Sudakov minoration and Gaussian concentration of measure, one can show that

$$\mathbb{P}(\hat{\pi} = \hat{\pi}^t) \leq e^{-cq(t)n^{1/2}},$$

where $q(t)$ is a function that behaves like t near zero and tends to a constant as $t \rightarrow \infty$.

- ▶ From this it follows that

$$\text{Var}(M_A) \leq n^{1/2} + n \int_0^\infty e^{-t} \mathbb{P}(\hat{\pi} = \hat{\pi}^t) dt \leq Cn^{1/2}.$$

- ▶ The above counterexample is also a counterexample to the notion that “hypercontractivity cannot go beyond $\log n$ improvements”.

- ▶ Let P_t be a Markov semigroup with invariant measure μ and generator L . Let λ_i and u_i be eigenvalues and eigenvectors of $-L$, as usual, with $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots$.
- ▶ We will say that a function f is ϵ -chaotic with respect to P_t if for all $t \geq \epsilon$,

$$e^{\lambda_1 t} \mathcal{E}(f, P_t f) \leq \epsilon \mathcal{E}(f, f).$$

When ϵ is small, we simply say that f is chaotic.

- ▶ We will see later that the left-hand side is a decreasing function of t , so it suffices to define ϵ -chaos as

$$e^{\lambda_1 \epsilon} \mathcal{E}(f, P_\epsilon f) \leq \epsilon \mathcal{E}(f, f).$$

Why do we call this chaos?

- ▶ In oriented last-passage percolation with Gaussian weights and O-U dynamics, $e^{\lambda_1 t} \mathcal{E}(L_N, P_t L_N)$ is precisely the expected number of edges in the intersection of the optimal paths at time 0 and at time t .
- ▶ This can be seen as follows:
 - ▶ Let \hat{Q}^t denote the optimal path at time t .
 - ▶ For the O-U semigroup, $\lambda_1 = 1$.
 - ▶ Recall that

$$\mathcal{E}(L_N, P_t L_N) = \mathbb{E}(\nabla L_N \cdot \nabla P_t L_N) = \sum_{e \in E(\mathbb{Z}^d)} \mathbb{E}[(\partial_e L_N)(\partial_e P_t L_N)],$$

where ∂_e is derivative with respect to ω_e .

- ▶ Recall that for any f , $\partial_e P_t f = e^{-t} P_t \partial_e f$.
- ▶ Lastly, recall that $\partial_e L_N = 1\{e \in \text{optimal path}\}$.
- ▶ Combining the above we get

$$e^{\lambda_1 t} \mathcal{E}(L_N, P_t L_N) = \sum_{e \in E(\mathbb{Z}^d)} \mathbb{P}(e \in \hat{Q}^0 \cap \hat{Q}^t) = \mathbb{E}|\hat{Q}^0 \cap \hat{Q}^t|.$$

Chaos (contd.)

- ▶ Similarly, in first-passage percolation with binary (integer) weights and Poisson flipping dynamics as the Markov process, $e^{\lambda_1 t} \mathcal{E}(T_N, P_t T_N)$ is the expected number of edges in the intersection of the pivotal sets at time 0 and at time t . (Pivotal set: Intersection of all optimal paths.)
- ▶ In both cases, ϵ -chaos with small ϵ means that within a short time of evolution, the new optimal path has very little intersection with the original optimal path.
- ▶ However, the environment does not change much in a short time of evolution. Thus, if chaos holds, then the optimal path depends very sensitively on the environment. This agrees with the usual notion of chaos. Indeed, the nomenclature is not my own; this phenomenon in disordered systems has been called chaos in numerous physics papers.
- ▶ The definition also coincides with the notion of “chaos in disorder” in the Sherrington-Kirkpatrick model.

Theorem (C., 2008)

If f is ϵ -chaotic, then it is δ -superconcentrated, where $\delta = \epsilon(\lambda_1 + 1)$. Conversely, if f is ϵ -superconcentrated, the f is δ -chaotic, where $\delta = \sqrt{\epsilon/\lambda_1} e^{\sqrt{\lambda_1}\epsilon}$.

Proof of equivalence

- ▶ Recall the formula

$$\mathcal{E}(f, P_t f) = \sum_{i=1}^{\infty} \lambda_i e^{-\lambda_i t} (u_i, f)^2.$$

- ▶ Clearly, this shows that both $\mathcal{E}(f, P_t f)$ and $e^{\lambda_1 t} \mathcal{E}(f, P_t f)$ are non-negative and decreasing functions of t .
- ▶ Suppose that f is ϵ -chaotic. Since $\mathcal{E}(f, P_t f)$ is decreasing in t ,

$$\begin{aligned} \text{Var}_{\mu}(f) &= \int_0^{\infty} \mathcal{E}(f, P_t f) dt \\ &\leq \epsilon \mathcal{E}(f, f) + \epsilon \int_{\epsilon}^{\infty} e^{-\lambda_1 t} \mathcal{E}(f, f) dt \\ &\leq \frac{\epsilon(\lambda_1 + 1)}{\lambda_1} \mathcal{E}(f, f). \end{aligned}$$

- ▶ Since $1/\lambda_1$ is the optimal Poincaré constant, this shows that f is $\epsilon(\lambda_1 + 1)$ -superconcentrated.

Proof of equivalence (contd.)

- ▶ Next, suppose that f is ϵ -superconcentrated.
- ▶ Since $\mathcal{E}(f, P_t f)$ is non-negative and decreasing, we have that for any $s > 0$,

$$\begin{aligned}\frac{\epsilon}{\lambda_1} \mathcal{E}(f, f) &\geq \text{Var}_\mu(f) \geq \int_0^s \mathcal{E}(f, P_t f) dt \\ &\geq s \mathcal{E}(f, P_s f).\end{aligned}$$

- ▶ Since $e^{\lambda_1 t} \mathcal{E}(f, P_t f)$ is decreasing, therefore for any $t \geq s$,

$$e^{\lambda_1 t} \mathcal{E}(f, P_t f) \leq e^{\lambda_1 s} \mathcal{E}(f, P_s f) \leq \frac{\epsilon e^{\lambda_1 s}}{\lambda_1 s} \mathcal{E}(f, f).$$

- ▶ Let $s = \sqrt{\epsilon/\lambda_1}$ and $\delta = se^{\sqrt{\lambda_1 \epsilon}} \geq s$. Then for any $t \geq \delta$,

$$e^{\lambda_1 t} \mathcal{E}(f, P_t f) \leq \frac{\epsilon e^{\lambda_1 s}}{\lambda_1 s} \mathcal{E}(f, f) = \delta \mathcal{E}(f, f).$$

This completes the proof.

- ▶ The equivalence between chaos and superconcentration led to proofs of chaos conjectures in first and last-passage percolation and the Sherrington-Kirkpatrick model. The conjectures were made by the physicists in the 80's.

For example:

Theorem (C., 2008)

Consider oriented last-passage percolation from $(0, \dots, 0)$ to (N, \dots, N) , evolving according to O-U dynamics. Let \hat{Q}^t be the optimal path at time t . Then for any $t \geq (\log N)^{-1/2}$,

$$\mathbb{E}|\hat{Q}^0 \cap \hat{Q}^t| \leq \frac{CN}{\sqrt{\log N}}.$$

Multiple peaks

- ▶ Chaos implies the existence of “multiple peaks”.
- ▶ In the context of last-passage percolation, this means simply the following: **With high probability, there exists a large number of paths that are nearly pairwise disjoint and are all nearly optimal.**
- ▶ Why would chaos imply such a thing?
 - ▶ Due to chaos, a small perturbation of the environment results in a new optimal path that is nearly disjoint from the old one.
 - ▶ Since the perturbation was small, the new optimal path must have been near-optimal in the old environment.
 - ▶ This establishes the existence of at least two near-optimal paths that are nearly disjoint. Repeating this procedure finds many such paths.
- ▶ The above sketch may be generalized to give a general theorem that says chaos (or equivalently, superconcentration) implies multiple peaks in the Gaussian setting.

A general formulation of multiple peaks for Gaussian fields

- ▶ Let $g = (g_i)_{i \in S}$ be a centered Gaussian field with index set S .
- ▶ Let $R(i, j) = \mathbb{E}(g_i g_j)$. Assume that $R(i, i) = 1$ and $R(i, j) \geq 0$ for all i, j .
- ▶ To fix ideas, think of S as the set of paths in oriented last-passage percolation with Gaussian edge-weights, and g_i as the weight of path i , divided by \sqrt{N} to have variance 1. Then $R(i, j)$ is the the number of common edges between paths i and j , divided by N .
- ▶ The ϵ -near-maximal region of g will refer to the random set

$$M_\epsilon(g) := \{i : g_i \geq (1 - \epsilon) \max_{j \in S} g_j\}.$$

- ▶ We will say that the field g has the ϵ -multiple peaks property if, with probability at least $1 - \epsilon$, there are at least $1/\epsilon$ points in $M_\epsilon(g)$, such that for any two of these points i and j , $R(i, j) \leq \epsilon$.

Chaos of the maximum implies multiple peaks in Gaussian fields

Theorem (C., 2008)

Suppose that the maximum of a Gaussian field g is ϵ -chaotic for some positive ϵ . Then g has the δ -multiple peaks property, where

$$\delta = \frac{C}{\sqrt{\log(1/\epsilon)}},$$

and C is a universal constant.

Implication for the S-K model

- ▶ Given i.i.d. standard Gaussian random variables $(g_{ij})_{1 \leq i, j \leq N}$, with high probability there exists a large number of nearly mutually orthogonal elements $\sigma \in \{-1, 1\}^n$ that nearly maximize the quadratic form

$$\sum_{i, j=1}^N g_{ij} \sigma_i \sigma_j.$$

- ▶ The same is **not true** for the linear form $\sum g_i \sigma_i$. Any near-maximal σ must be nearly equal to the optimizer.
- ▶ Intuitively, why this difference between the quadratic form and the linear form? I don't know.

Chaos under large perturbations implies chaos under small perturbations

- ▶ Take any f and let

$$\phi(t) := e^{\lambda_1 t} \mathcal{E}(f, P_t f).$$

- ▶ Say that f is ϵ -chaotic at time t if $\phi(t) \leq \epsilon \mathcal{E}(f, f)$. (So that f is ϵ -chaotic if f is ϵ -chaotic at time ϵ .)
- ▶ Recall the formula

$$\phi(t) = \sum_{i=1}^{\infty} \lambda_i e^{-(\lambda_i - \lambda_1)t} (u_i, f)^2.$$

- ▶ An easy verification using this shows that ϕ is log-convex.
- ▶ Consequently, for any $0 < t \leq s$,

$$\phi(t) \leq \phi(0)^{1-t/s} \phi(s)^{t/s} = \mathcal{E}(f, f)^{1-t/s} \phi(s)^{t/s}.$$

- ▶ Thus, if $\phi(s) \leq \epsilon \mathcal{E}(f, f)$, then $\phi(t) \leq \epsilon^{t/s} \mathcal{E}(f, f)$. In other words, ϵ -chaos at time s implies $\epsilon^{t/s}$ -chaos at time t .

Application to the S-K model

- ▶ Let $f = F_N(\beta)$ be the free energy of the S-K model with N particles at inverse temperature β .
- ▶ Let ϕ be defined as in the previous slide.
- ▶ Then one can define an infinite sequence of functions ϕ_0, ϕ_1, \dots , with $\phi_0 = \phi$, such that for any k and t ,

$$0 \geq \phi'_k(t) \geq -4\beta^2 e^{-2t} \phi_{k+1}(t).$$

- ▶ This chain of differential inequalities and information about the behavior of the ϕ_k 's as $t \rightarrow \infty$ may be combined to deduce that there exists $s > 0$ independent of N such that

$$\phi(s) \leq \frac{C(\beta)}{N} \mathcal{E}(f, f).$$

In other words, f is $C(\beta)N^{-1}$ -chaotic at time s .

Application to the S-K model (contd.)

- ▶ Therefore, f is $C(\beta)N^{-t/s}$ -chaotic at any time $t \leq s$.
- ▶ This information, plugged into the semigroup method, shows that the variance of the free energy is bounded by $C(\beta)N/\log N$, improving our earlier bound by removing the extra $\log \log N$ factor.

Extremal Gaussian fields

- ▶ Let $g = (g_1, \dots, g_n)$ be a centered Gaussian random vector, whose components may be correlated. (Gaussian vector, Gaussian field, same thing.)
- ▶ Let $R(i, j) = \mathbb{E}(g_i g_j)$. Assume that $R(i, i) = 1$ for all i .
- ▶ We say that g is **extremal** if $\mathbb{E}(\max g_i)$ is close to $\sqrt{2 \log n}$.
- ▶ Since this is an imprecise definition, we may make it precise by saying that g is ϵ -extremal if

$$\mathbb{E}(\max g_i) \geq (1 - \epsilon) \sqrt{2 \log n}.$$

- ▶ We do not need an upper bound in the definition, because it can be proven that $\mathbb{E}(\max g_i) \leq \sqrt{2 \log n}$ without any further assumptions.
- ▶ Often in the presence of low correlations we can prove extremality using Slepian type comparison inequalities for expectation of maxima, or other methods (vast theory).
- ▶ Several interesting Gaussian fields are extremal in spite of the presence of strong correlations.

Theorem (C., 2008)

If g is ϵ -extremal, then the maximum of g is $\delta(\epsilon, n)$ -superconcentrated, where $\delta(\epsilon, n)$ is a function of ϵ and n such that

$$\lim_{(\epsilon, n) \rightarrow (0, 0)} \delta(\epsilon, n) = 0.$$

- ▶ This theorem helped establish, for the first time, the superconcentration of the maximum of the two-dimensional **discrete Gaussian free field** (DGFF).
- ▶ Used an earlier result of Bolthausen, Deuschel and Giacomin who proved that the DGFF is an extremal field.
- ▶ The DGFF superconcentration result has been vastly refined recently by Bramson and Zeitouni, and even more recently by Ding.

Some open questions

- ▶ Improve the variance bounds in any of the problems mentioned: the S-K model free energy, last-passage percolation, first-passage percolation. Optimal bounds would be best, but even small improvements would be remarkable.
- ▶ Provide a definition of multiple peaks that makes it equivalent to chaos and superconcentration.
- ▶ Prove superconcentration for the ground state energy of the S-K model (i.e. $F_N(\infty)$). I can only do this for the free energy at non-zero temperature.
- ▶ Establish chaos in temperature in the S-K model (i.e. chaotic behavior under small perturbations of β).
- ▶ Improve the multiple peaks theorem. Physicists say that there are exponentially many peaks, while the theorem gives only a logarithmic number of them.